DOI:
10.1109/ARES.2010.42

Document Version
Peer reviewed version

A Complexity Based Model for Quantifying Forensic Evidential Probabilities

Richard E Overill and Jantje A M Silomon

Department of Computer Science, King's College London, Strand, London WC2R 2LS

1. Introduction and Background

The basis of an operational complexity model (OCM) is presented. The model is based on the observation that an inverse relationship holds between the difficulty and/or intricacy involved in performing a task in a specified manner, as measured by its intrinsic complexity, and the likelihood that the task in question was in fact performed in the specified manner.

Thus, in the context of a digital forensic examination, the operational complexity of formation of a set of digital evidential traces by a specified route should in principle be susceptible to 'bottom-up' *ab initio* determination. The resultant complexity should then be inversely related to the probability of formation by that route.

There are many definitions of complexity. Lloyd [1] lists several complexity measures which in principle permit the complexity of formation of a set of digital evidential traces $\{E_i\}$ to be defined. These metrics include computational complexity, information based complexity, logical depth, thermodynamic depth and crypticity. The data available in the problem space of digital forensic analysis appears to be most closely addressed by the computational complexity metric. In addition, it is desirable to include in the model a component relating to the human (or cognitive) complexity of the task. The GOMS (Goals, Operators, Methods, Selections) family of models offers a well-understood approach to the problem. In particular, the GOMS Keyboard-Level Model (KLM) [2] provides a tractable means of measuring the human involvement in the operational process.

This development offers forensic examiners and expert witnesses the possibility of computing the probabilities that a given set of recovered digital evidential traces was formed via a number of alternative (mutually exclusive) routes.

2. The Operational Complexity Model

The resulting model may be formalized as follows. The various ***feasible routes*** by which the recovered set of digital evidential traces could have been formed are first enumerated. For each feasible route k by which the set of digital evidential traces $\{E_i\}$ could have been formed the operational complexity of that route is given by:

$$C_k = KLM_k + CC_k$$

where C_k comprises a cognitive complexity component specified by the GOMS-KLM and a suitably defined computational complexity (CC) component.

The operational complexity of each feasible route C_k and its probability of occurrence p_k are inversely related:

$$p_k \propto C_k^{-1}$$

The constant of proportionality is determined uniquely by the normalization condition on the probabilities:

$$\sum p_k = 1$$

The constant of proportionality α reflects the units in which the complexity of each of the feasible routes k is measured, and is given by:

$$\alpha = (\sum C_k^{-1})^{-1}$$

It should be noted here that while the OCM model makes use of a complexity metric it is not based on Shannon information theory, which would lead to an inverse exponential relation:

$$p_k \propto 2^{-Ck}$$

In the context of Bayesian network models, the ***posterior probability*** of a feasible route k to the formation of a recovered set of digital evidential traces $\{E_i\}$ is given by $\Pr(H_k|\{E_i\})$ where H_k represents the hypothesis that feasible route k was taken. The ***odds*** for two (mutually exclusive) alternative routes k and k' to the formation of the recovered set of digital evidential traces $\{Ei\}$ is then given by:

$$O(k{:}k') = \Pr(H_k|\{E_i\}) \,/\, \Pr(H_{k'}|\{E_i\})$$

In a digital forensics context, if H_k represents the prosecution's contention regarding the formation of $\{Ei\}$ and $H_{k'}$ is the defence's alternative contention, then the odds $O(k{:}k')$ provides a valuable measure of the relative plausibility of the two competing narratives.

3. Application to the BitTorrent Case

To illustrate the use of the operational complexity model outlined above we give here an application to the BitTorrent case described previously [3]. The prosecution case is taken to be exactly as described in [3]. The defence's alternative explanation for the presence of the recovered set of digital evidential traces $\{Ei\}$ is assumed to be as

follows. A Trojan horse carrying the multimedia file as part of its payload installed itself on the defendant's computer and invoked the µTorrent client to upload the multimedia file to a peer-to-peer (P2P) file sharing website, before finally uninstalling itself.

We have made a number of simplifying assumptions for the purposes of this illustration. The Trojan horse is not equipped with its own life-support system; the computer is not protected by a firewall or an anti-malware scanner. The basic unit of the KLM characterization [2] of cognitive information processing is taken to be the mouse button press or release; similarly, the basic unit of information processing used in characterizing the computational complexity is the byte. Disk accesses are assumed to take place autonomously and concurrently with CPU- and RAM-based processes. Given these assumptions and using documented or typical values for all other quantities (see the Appendix for full details) we obtain the following results:

$KLM_k = 510$

$KLM_{k'} = 0$

$CC_k = N + 20N/2^{19} + 1{,}844{,}346$

$CC_{k'} = (23/5)N + 20N/2^{19} + 9{,}938{,}941$

Taking a typical value for the size of the multimedia file as N = 4GB, we obtain:

$CC_k = 4{,}296{,}975{,}482$

$CC_{k'} = 19{,}766{,}952{,}343$

Hence, providing that route *k'* is the ***only*** feasible alternative to route *k,* we find $O(k{:}k') \approx 4.60$, indicating that the prosecution's case is 4.6 times more plausible than the defence's case, given the recovered evidence. Alternatively, in the absence of any other feasible explanations, the probability that the prosecution's case is correct is ≈ 82%.

4 Summary and Conclusions

The recently developed operational complexity model enables the complexity of both the cognitive and the computational components of a process to be determined. From the complexity of formation of a set of traces via a specified route the probability of that route can be determined. By determining the complexities of alternative routes leading to the formation of the same set of traces, the odds indicating the relative plausibility of the alternative narratives can be found. An illustrative application to the previously discussed BitTorrent case has been presented, and the results obtained suggest that the proposed operational complexity model is capable of providing a realistic estimate of the odds for two competing hypotheses. This finding should prove useful for forensic examiners and expert witnesses as they seek to evaluate the strength of a case given the recovered digital evidence.

References

[1] S Lloyd, Measures of Complexity: a Nonexhaustive List, *IEEE Control Systems*, **21 (4)** August 2001, 7-8.

[2] D Kieras, Using the Keystroke-Level Model to Estimate Execution Times, University of Michigan (2001), available online at http://www.cs.loyola.edu/~lawrie/CS774/S06/homework/klm.pdf

[3] M. Kwan, K.P. Chow, F. Law & P. Lai. Reasoning About Evidence using Bayesian Network, *Advances in Digital Forensics IV*, International Federation for Information Processing (IFIP) January 2008, Tokyo, pp.141-155.

Appendix

KLM										
				Action	M	P	B	K	H	Total
K (key press & release)	2									
			1	Drag and Drop	2	2	2	0	0	48
P(point the mouse)	11		2	Double click	1	1	4	0	0	27
			3	Single click	1	1	2	0	0	25
B(button press/release)	1		4	Create torrent	5	6	10	0	0	136
			5	Upload torrent	5	5	10	0	0	125
H (hand to/from keyboard)	4		6	Type URL	2	1	4	16	2	79
			7	Log in (username/pw)	4	2	4	16	4	122
M (mental preparation)	12									

Variables	Description	BT specific values
N	no. of data bytes in file to be shared	4GB
N_{THC}	no. of data bytes of Trojan Horse Code	128KB
N_{THD}	no. of data bytes in Trojan Horse Dropper program	$(N+N_{THC})/IFL$
N_{TC}	no. of data bytes in Torrent Client	7MB
N_{TCI}	no. of data bytes in Torrent Client Installation file	276KB
N_{DI}	no. of data bytes in Desktop.ini file	47B
TD	TimeDate read or write	8B
TFN	Torrent File Name	256B
TFL	Torrent File Length	4B
TPL	Torrent Piece Length	4B
TPS	Torrent PieceSize	512KB
TAU	Tracker Announce URL	35B
IFL	Inflation factor (unzip)	1.25
DSK	Disk access (assumed autonomous)	0
PCI	Peer Connection Information process	52B + 3TD + DSK
TSL	Tracker Server Login process	60B + 3TD + DSK
PG	Page creation process (webpage)	600KB + TD + DSK
CO	Cookie creation process	256B + TD + DSK
CA	Cache creation process	16B + TD + DSK

Route *k* (Criminal)			
Evidence (see [4])	KLM actions	DSK	Bytes
E1 (incl. E2, 3, 5, 6)	1	3	N + 7TD
E4	0	0	0
E7 (incl. E8)	2+4	5	20N/TPS + TFN + TFL + TPL + TAU + 11TD
E9 (incl. E10-12, 14, 15, 17, 18)	2+3+5+7	11	3PG + CO + 3CA + N_{DI} + PCI + TSL + 17TD
E13 & E16	0	0	0

Route *k'* (Trojan)			
Evidence (see [4])	KLM actions	DSK	Bytes
DSI (Dropper S/W Install)	0	1	N_{THD} + 3TD
DSU (Dropper S/W Uninstall)	0	1	N_{THD} + 3TD
TSI (Trojan S/W Install & payload copy incl. E1, 2, 3, 5, 6)	0	4	N_{THD}*IFL + N + 10TD
E4	0	1	16 + N_{TCI} + N_{TC} + 3TD
E7 (incl. E8)	0	2	20N/TPS + TFN + TFL + TPL + TAU + 4TD
E9 (incl. E10-12, 14, 15, 17, 18)	0	11	3PG + CO + 3CA + N_{DI} + PCI + TSL + 17TD
E13 & E16	0	0	0
TSU (Trojan S/W Uninstall)	0	1	N_{THD}*IFL + 3TD